Workshop on
The Future of the UMLS Semantic Network
NLM, April 8, 2005

Consistency between Metathesaurus and Semantic Network

Olivier Bodenreider

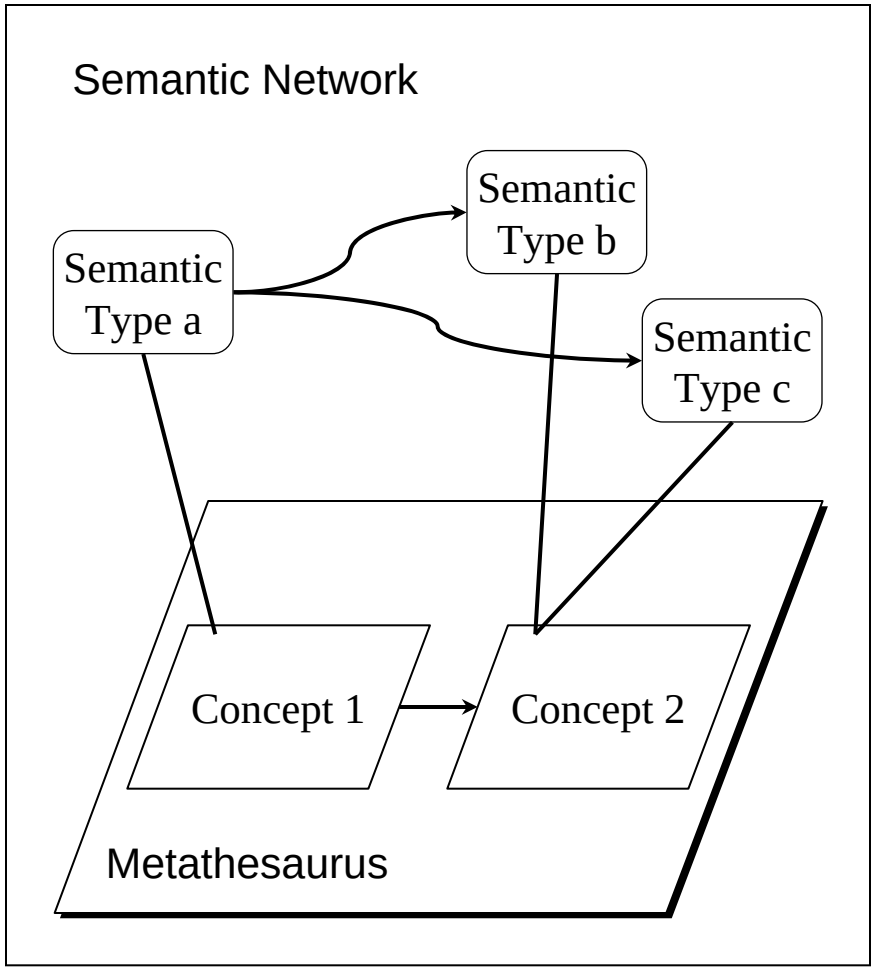
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

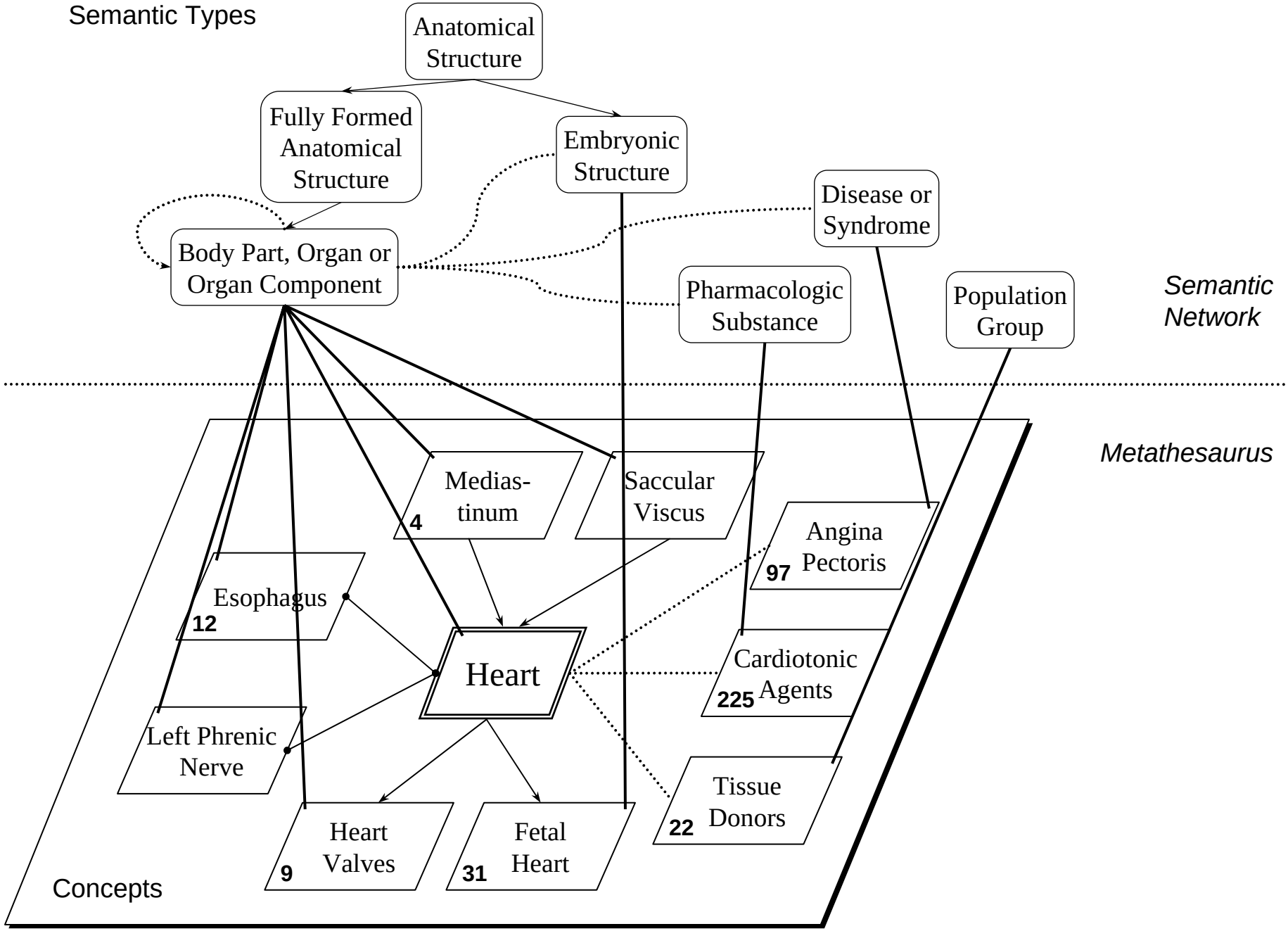
Overview

- Defining consistency
- What does inconsistency mean?
- Testing consistency
  - Comparing Metathesaurus relations to SN relations
  - Aligning Metathesaurus concepts and semantic types
  - Semantic type distribution of sets of descendants of Metathesaurus concepts
- Suggestions
  - Enforcement mechanism
  - Ontology of relationships
  - CVF

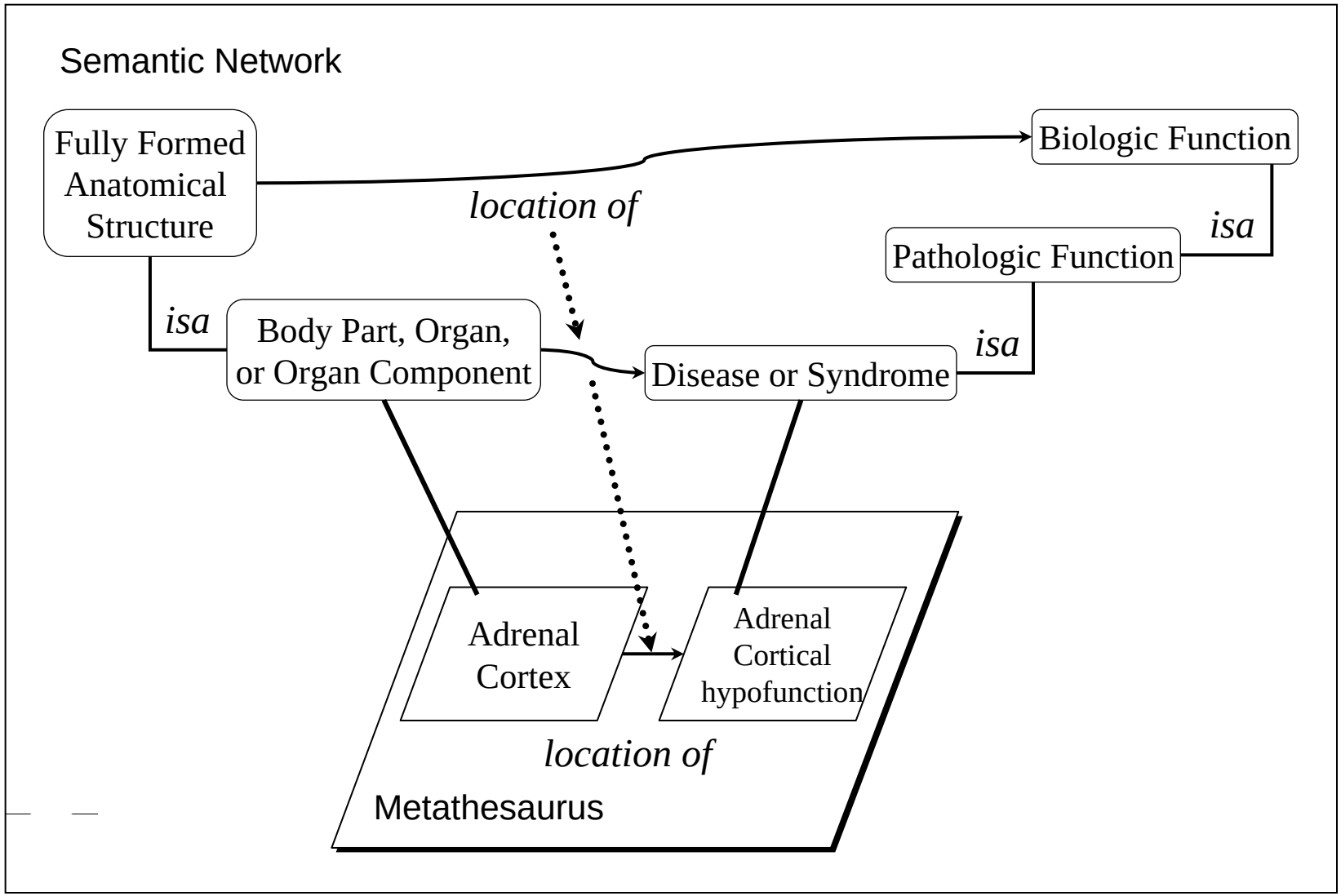
Two levels in the UMLS

The UMLS: a two-level structure



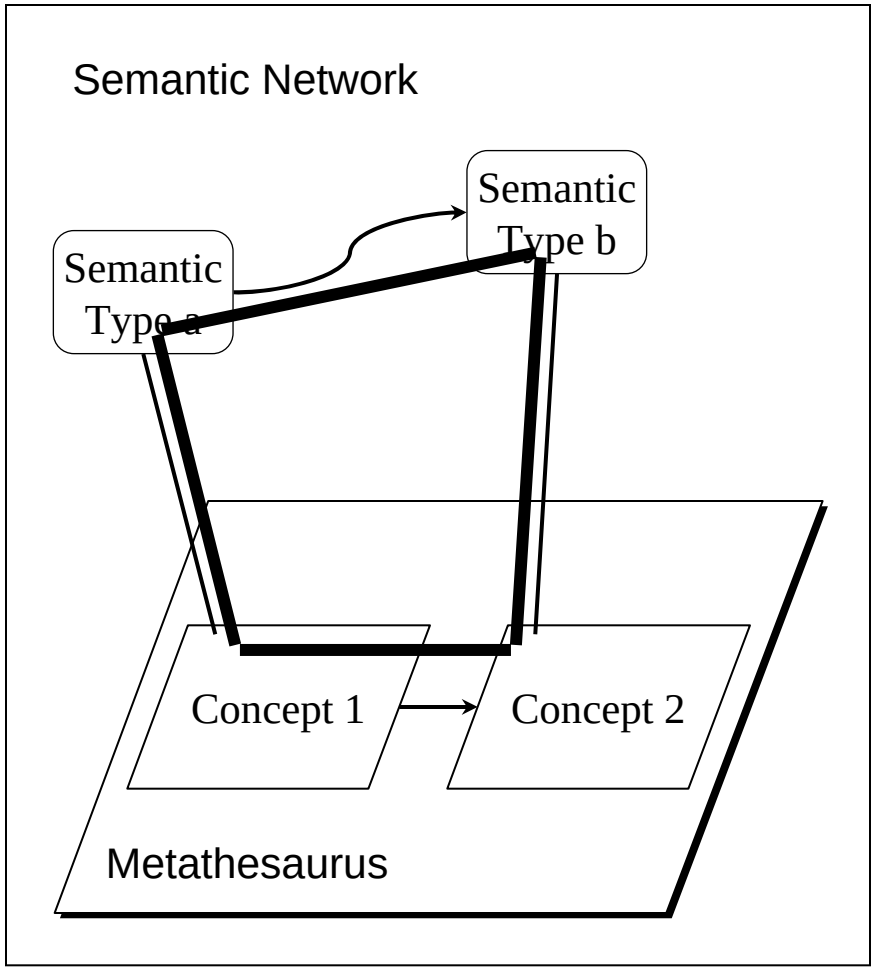

Semantic Types
Anatomical Structure
Fully Formed Anatomical Structure
Embryonic Structure
Disease or Syndrome
Body Part, Organ or Organ Component
Pharmacologic Substance
Population Group
Semantic Network
Metathesaurus
Medias-tinum
4
Saccular Viscus
Angina Pectoris
97
Esophagus
12
Heart
Cardiotonic Agents
225
Left Phrenic Nerve
Tissue Donors
22
Heart Valves
9
Fetal Heart
31
Concepts

Relationships can inherit semantics

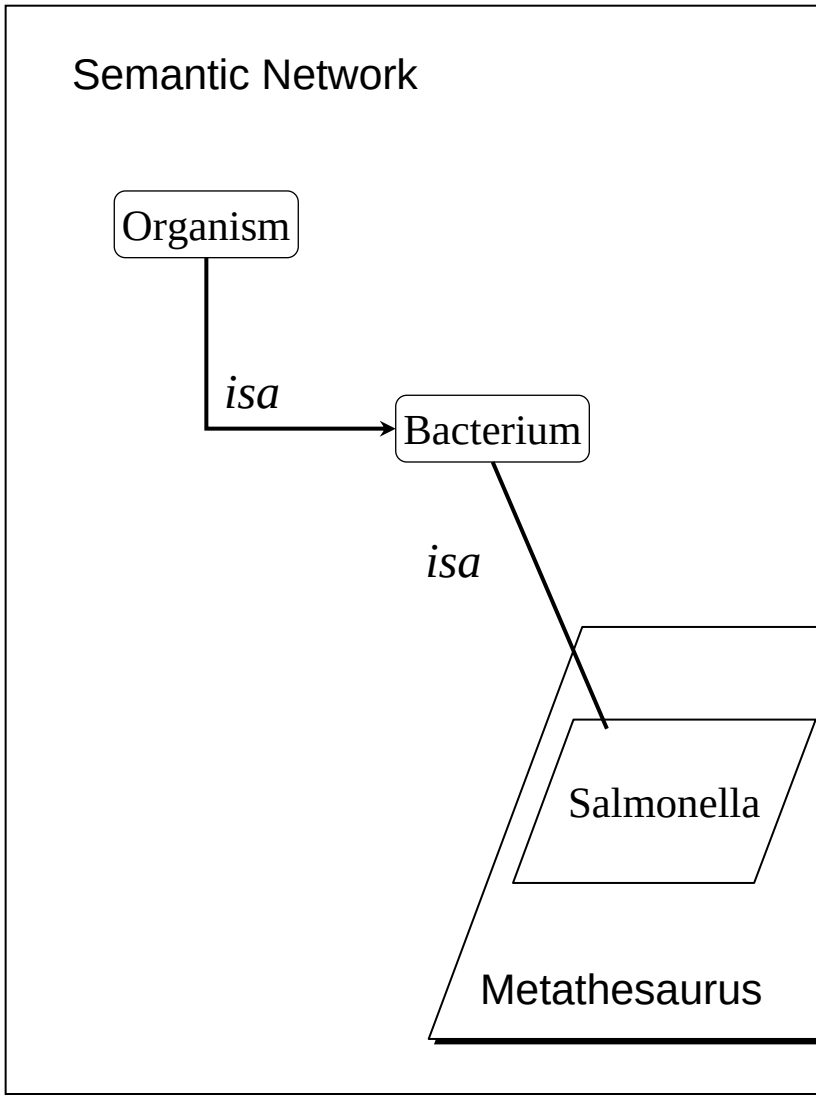


Defining consistency

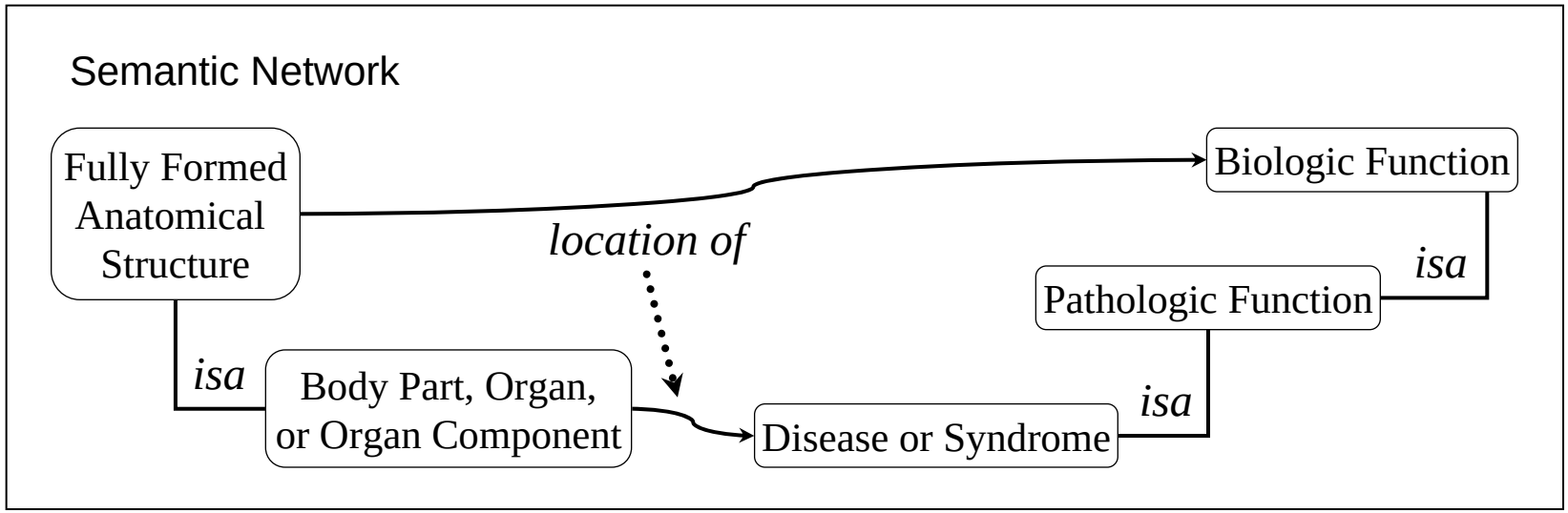
The consistency “square”



The categorization link



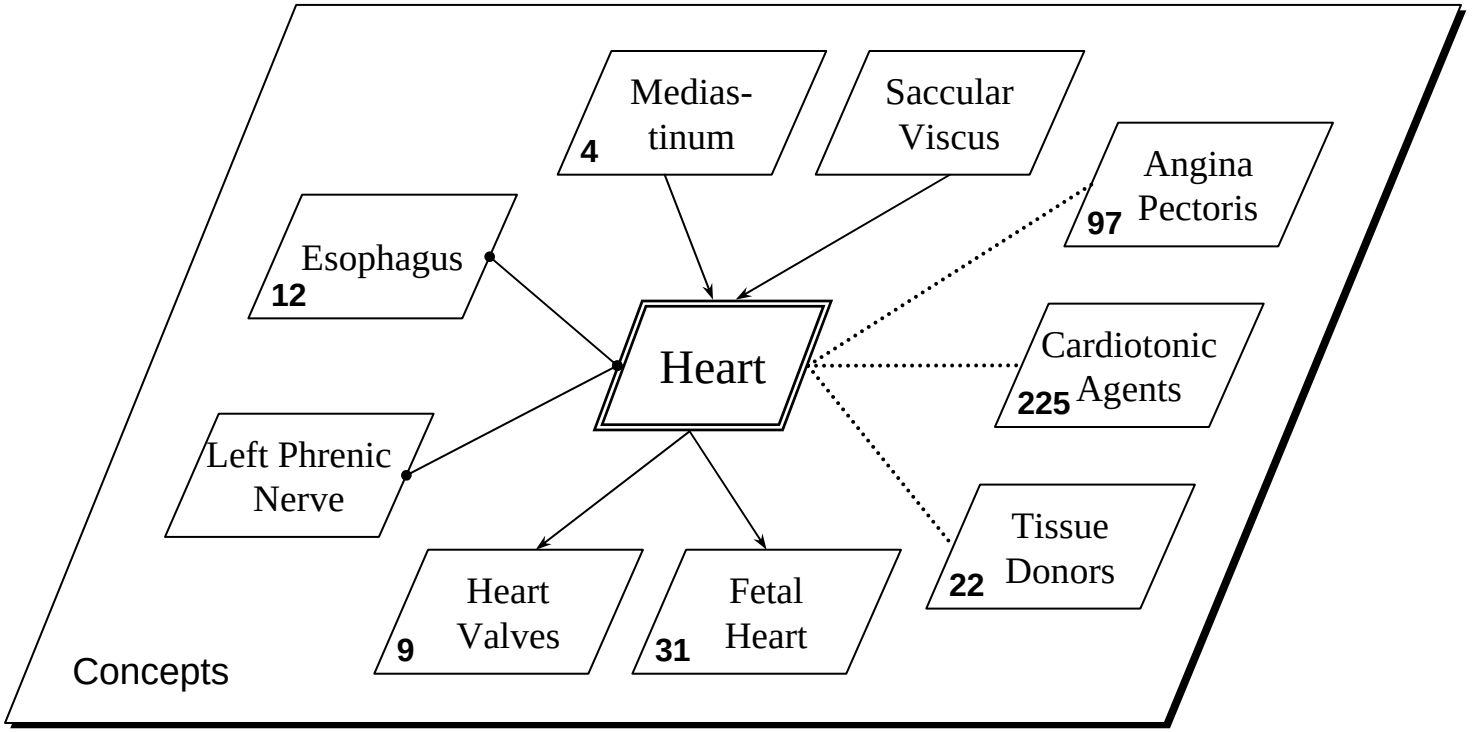
Semantic Network relations



- 54 types of relationships
- 558 asserted relations (SRSTR)
- 6703 fully expanded relations (SRSTRE*)

Metathesaurus relations

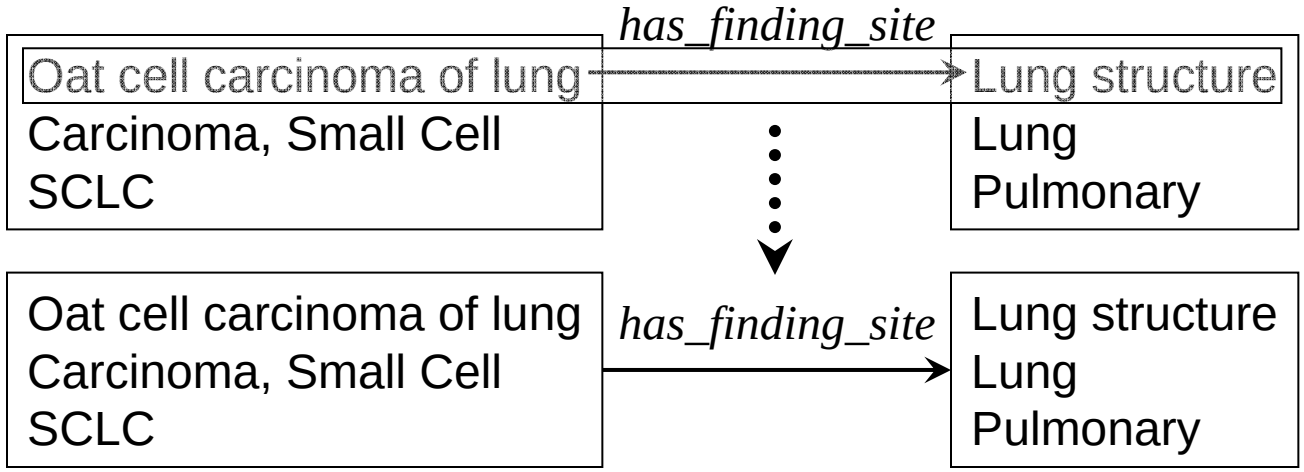
- REL vs. RELA
- Not always labeled
- 106 additional types of relationships
- ~7 M symbolic relations



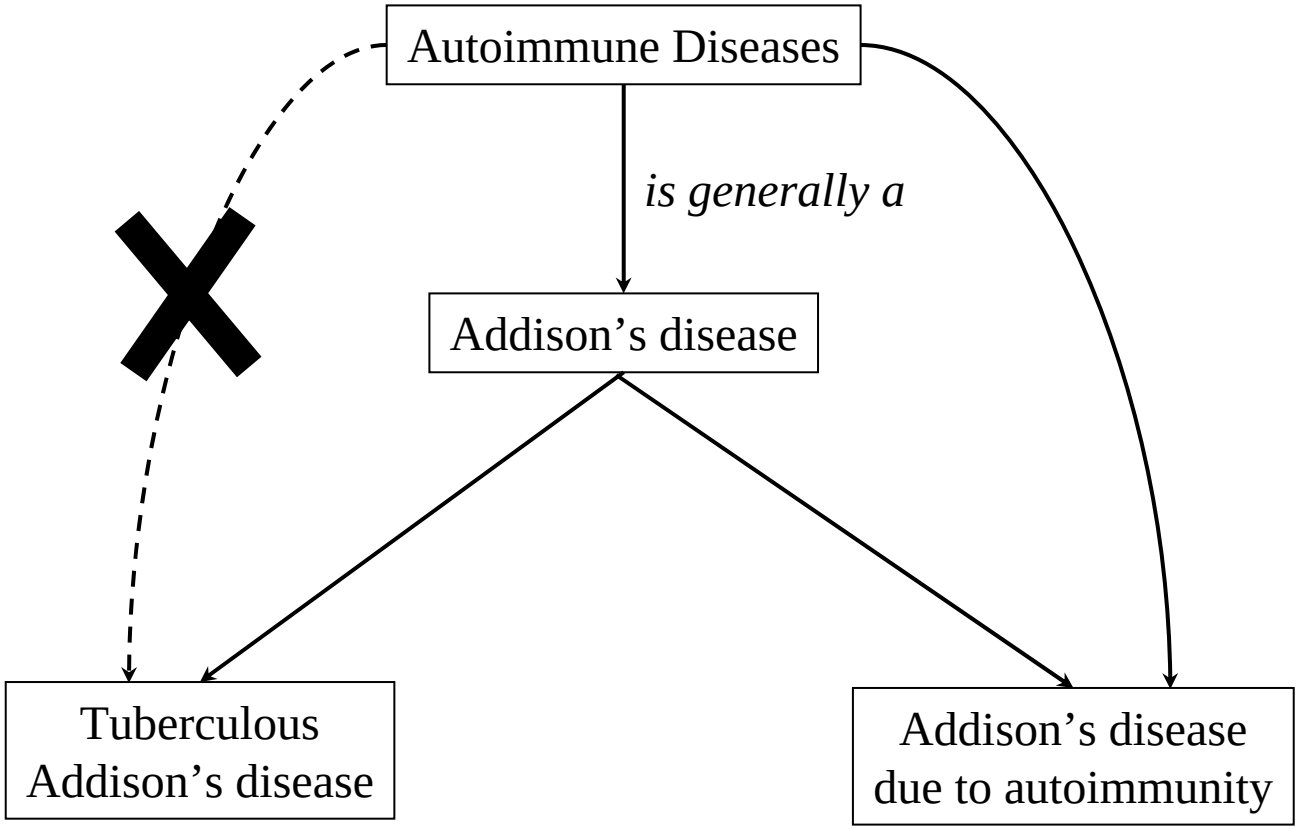
Metathesaurus

Metathesaurus relations

- Recorded
  - at the term level: from source vocabularies
  - at the concept level: from Metathesaurus editors
- Aggregated at the concept level



Not all relationships in hierarchies are *isa* (1)



Not all relationships in hierarchies are *isa* (2)

Environment and Public Health [G03]

Public Health [G03.850]

Accidents [G03.850.110]

Accident Prevention [G03.850.110.060] +

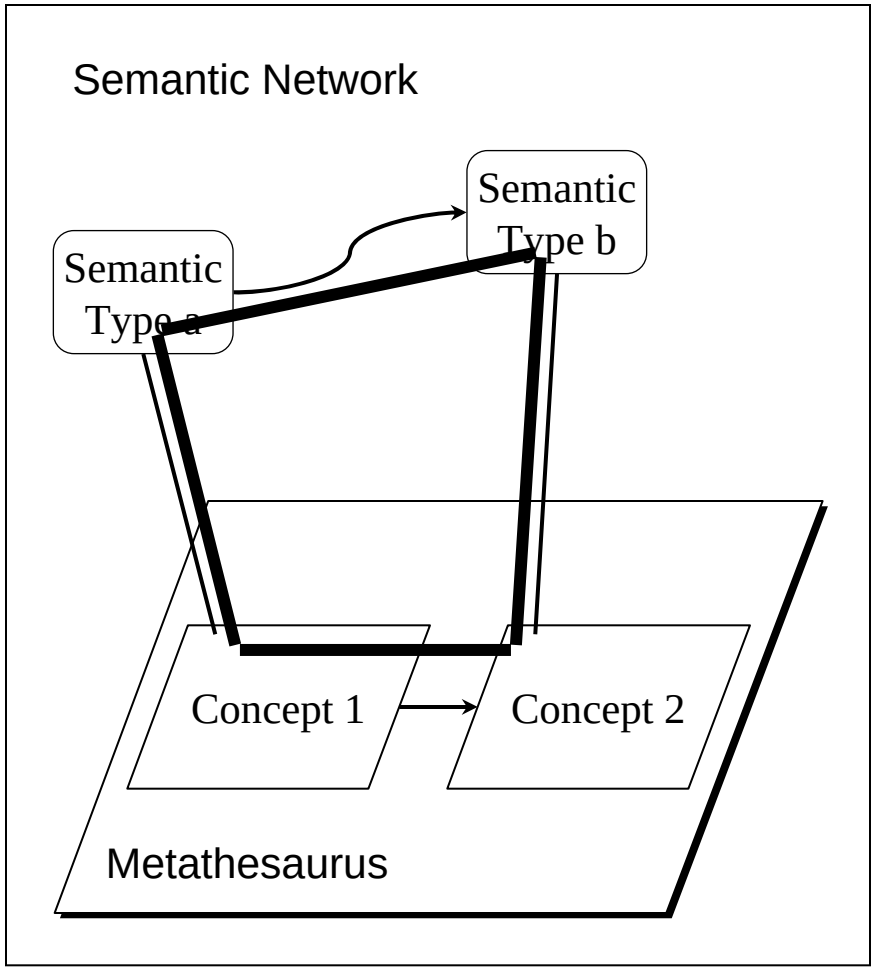
Accidental Falls [G03.850.110.085]

Accidents, Aviation [G03.850.110.185]

[…]

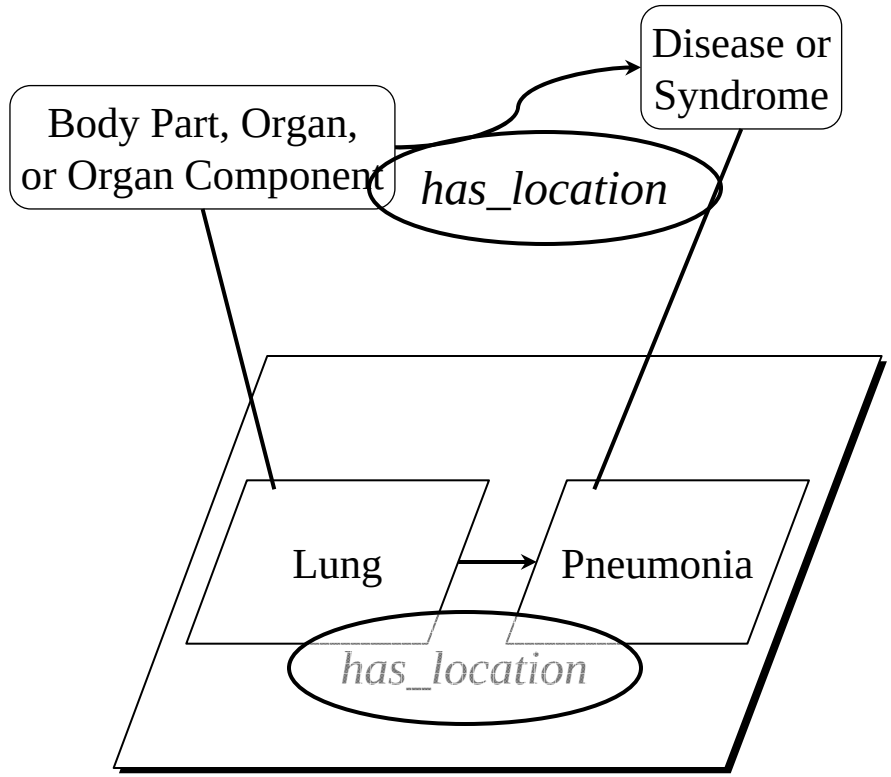
Drowning [G03.850.110.500] +

Defining consistency

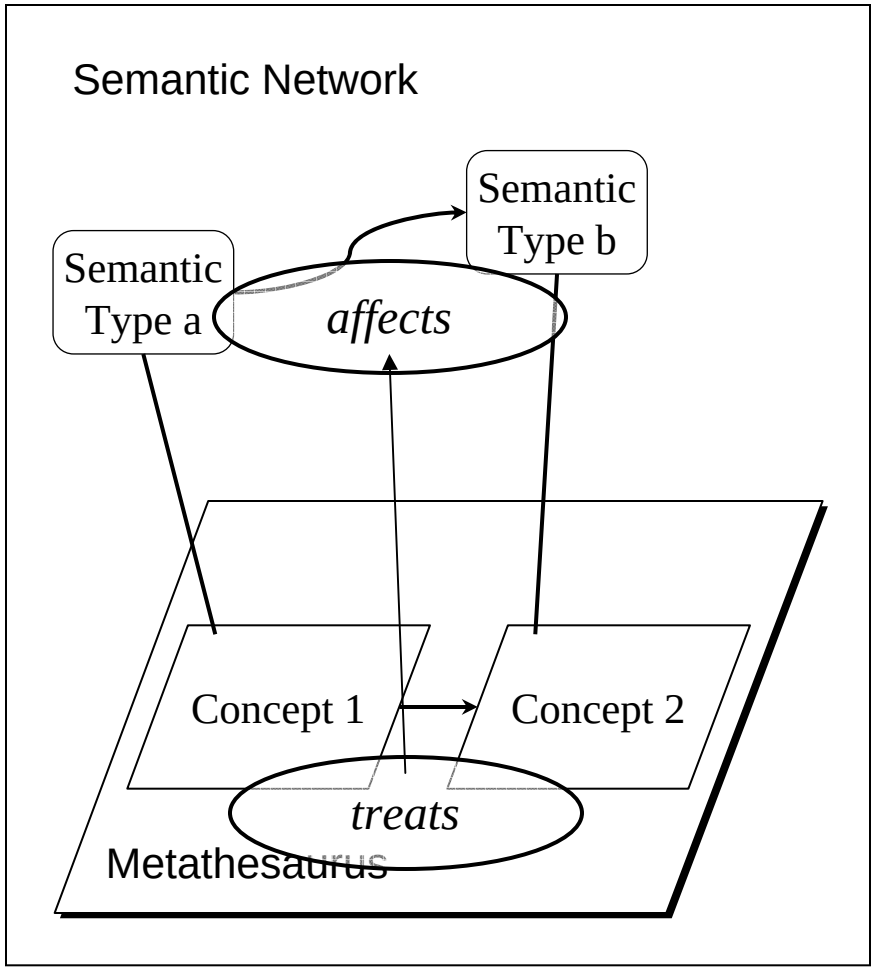


- SN rel. and Meta rel. must have the same direction
- SN rel. and Meta rel. must be of the same type (both hierarchical or associative)
- Meta rel. must be the same as SN rel. or one of its descendants

Examples of consistent relations

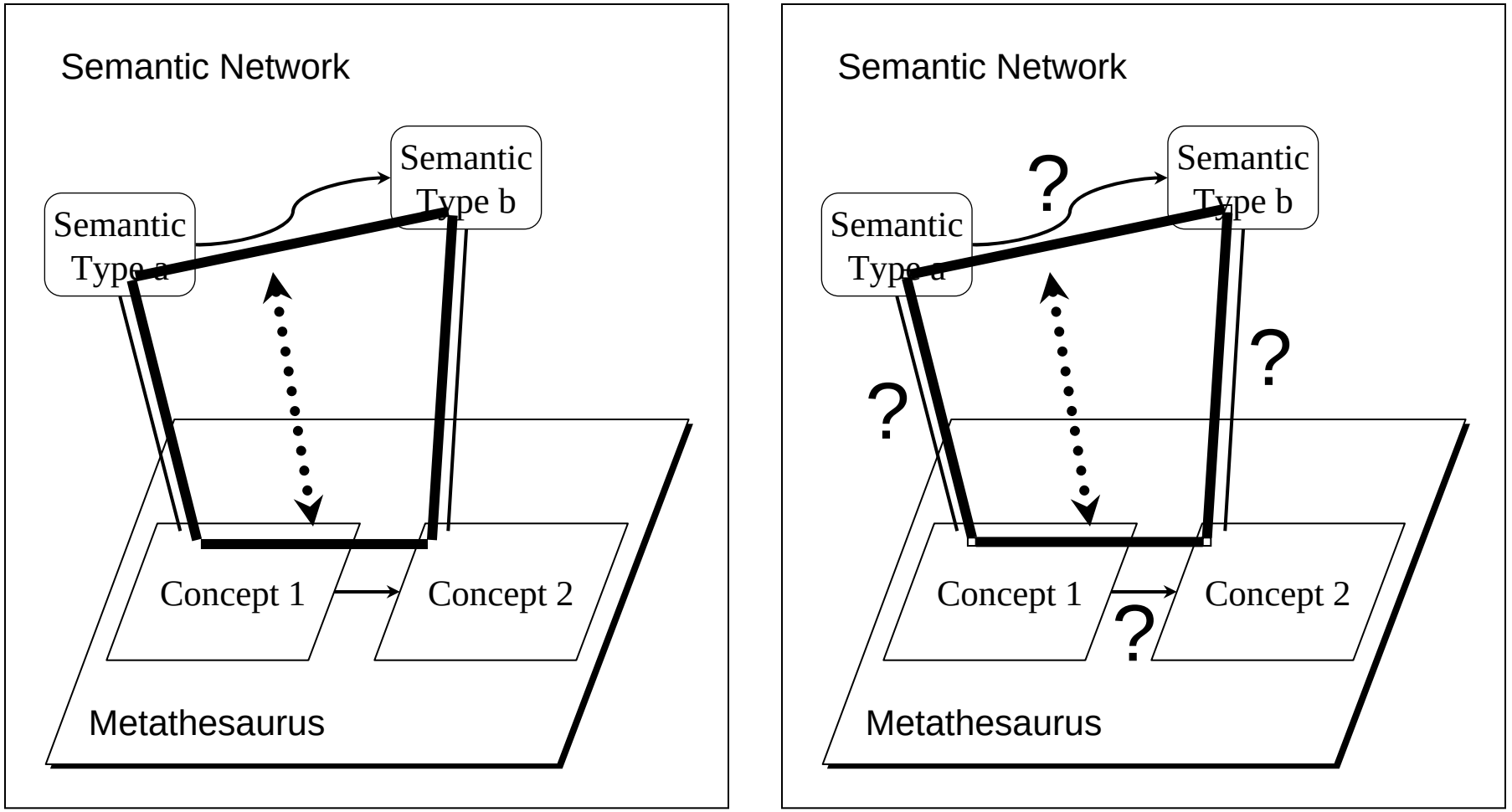


Examples of consistent relations



What does inconsistency mean?

The consistency “square” revisited

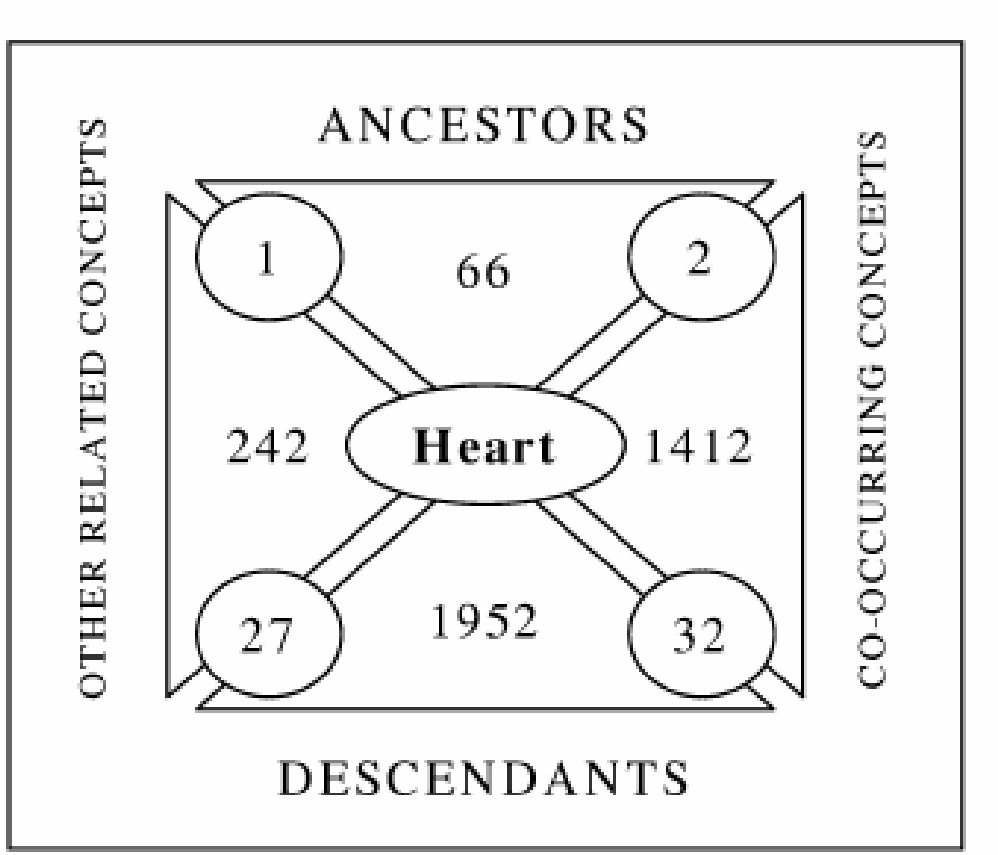


What does inconsistency mean?

- Inaccurate/missing Semantic Network relation
- Inaccurate (/missing?) categorization
- Inaccurate Metathesaurus relation

Testing consistency

(A) Consistency of associative relations



[McCray
& Bodenreider, 2002]

Results

- 6894 pairs of related concepts
  - 4496 (65%): a SN relation can be inferred unambiguously
    - Validity confirmed in 1981 cases
    - 2515 not labeled in the Metathesaurus
  - 1491 (22%): multiple possible SN relationships
    - multiple possible Metathesaurus relationships
  - 907 (13%): inconsistency SN/Meta relationships
    - 372: no SN relationship between the STs
    - 415: inconsistent SN/Meta relationship type (REL)
    - 120: inconsistent SN/Meta relationship attribute (RELA)

(B) Consistency of hierarchical relations

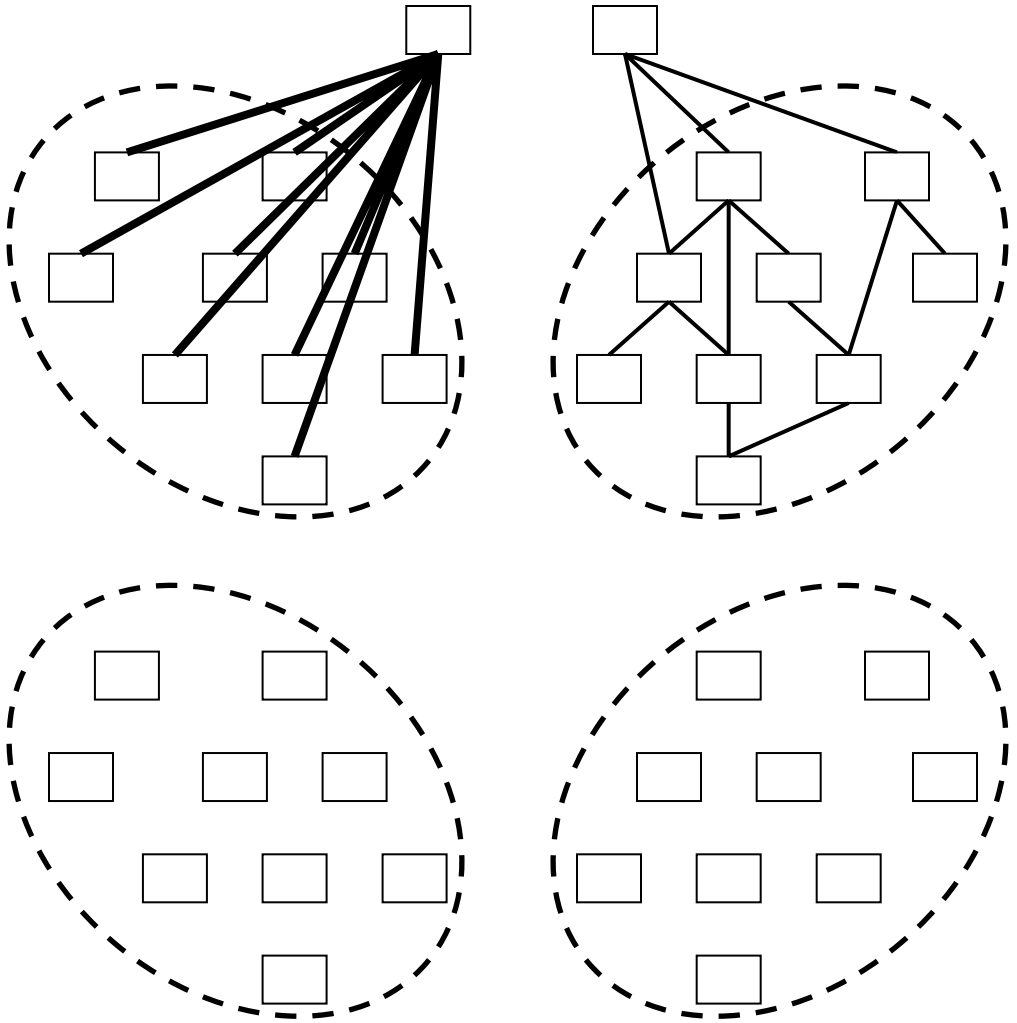
- Relations used
  - SN: isa
  - Categorization: isa
  - Metathesaurus: PAR/CHD + RB/RN
- Hypothesis

❶ - For a pair of (ST, C), the concepts categorized by ST (and its descendants) correspond to the descendants of the concept C

❷ - In the set of descendants of C, expected STs are the ST of C (and its descendants)

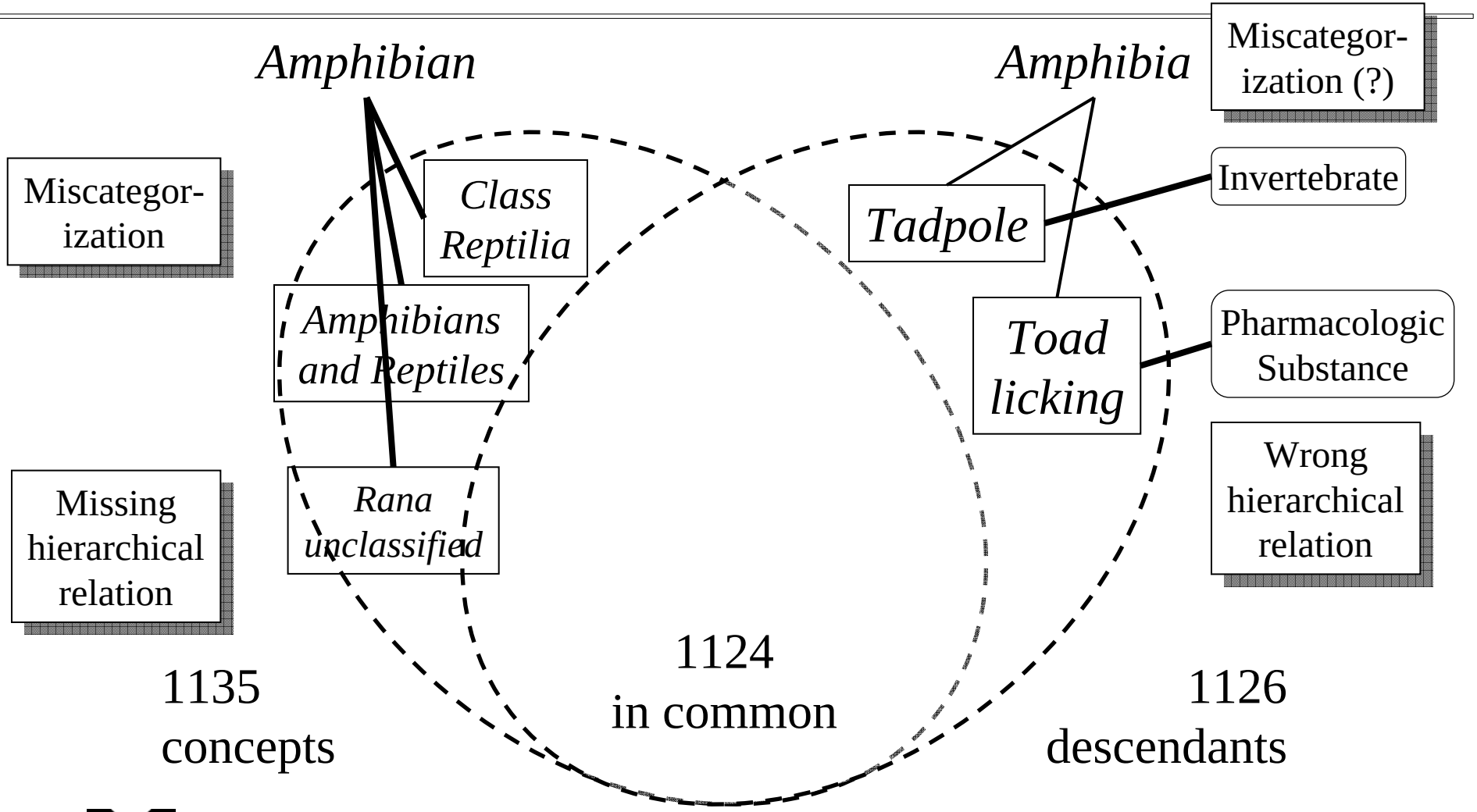
❶ ST-based classes vs. descendants

- Semantic type
  - List of all concepts having this semantic type
- Concept
  - List of all descendants
- Comparing the 2 sets
  - Intersection of the 2 sets



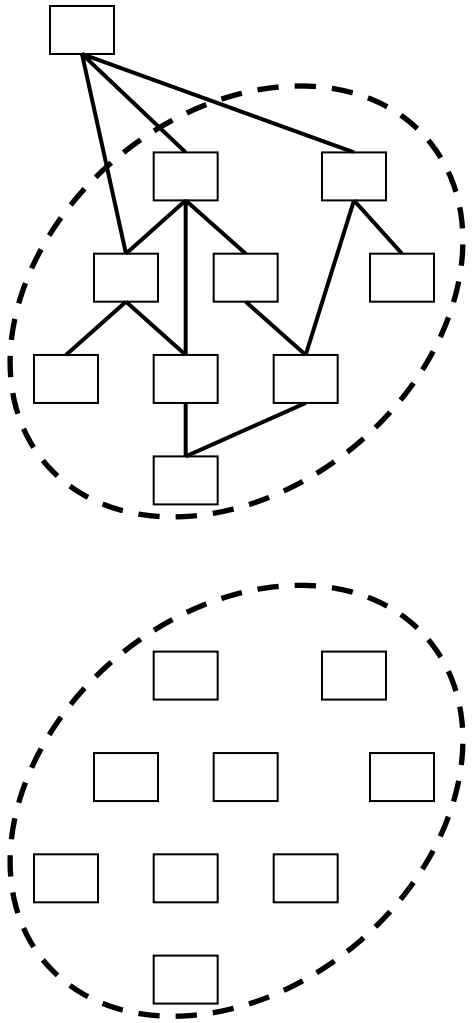
[Bodenreider & Burgun, 2004]

Analyzing inconsistencies



❷ Semantic types of descendants

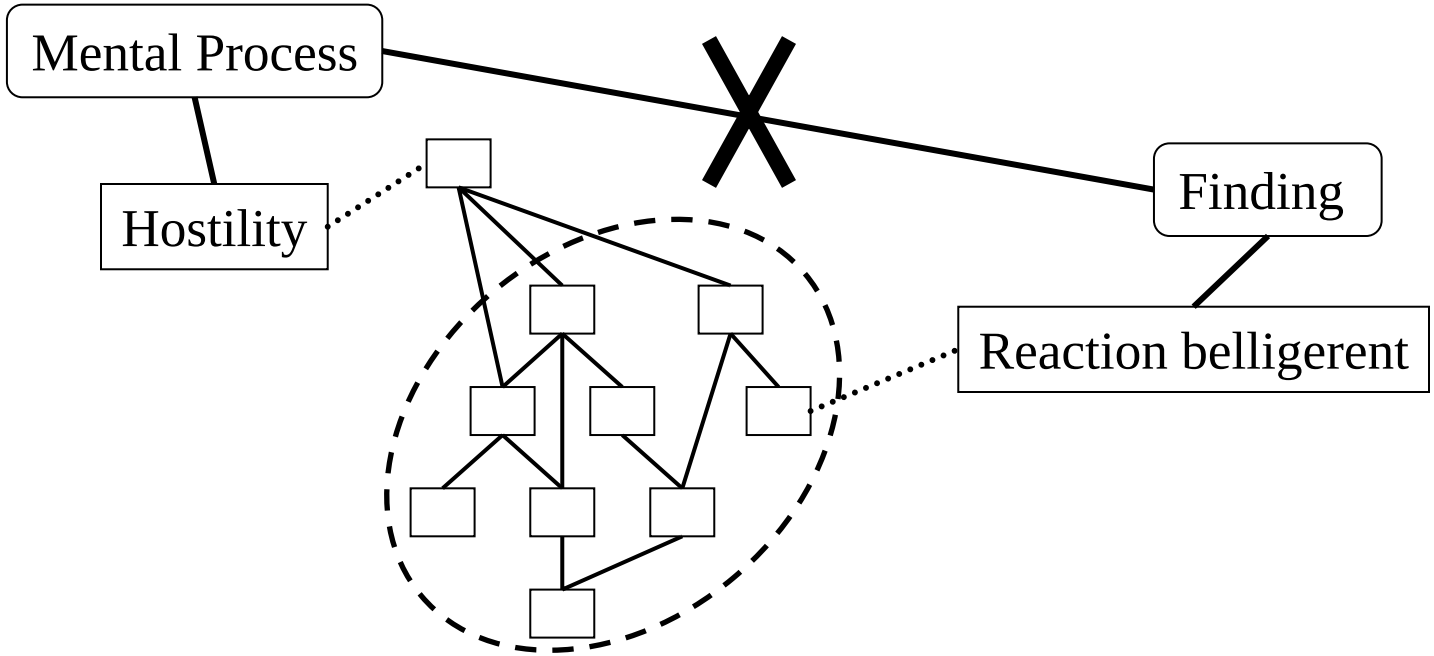
- Concept
  - Set of all descendants
- Distribution of semantic types in the set
  - Allowable STs: ST of C and its descendants (strict) or ST from the same semantic group (loose)



[Mougin & Bodenreider, 2005]

Analyzing inconsistencies

- 26,584 concepts studied
- 59% of their descendants have a semantic type incompatible with that of the original concept



Analyzing inconsistencies

```
# ------------------------------------------------------
# C0597249 Neoplasm of placenta (disorder) (neop)
# * B: 190

C0597249|ST|acab|    5.50|incp
C0597249|ST|anab|    1.50|incp
C0597249|ST|cgab|   76.50|incp
C0597249|ST|dsyn|   27.50|incp
C0597249|ST|inpo|    1.00|incp
C0597249|ST|neop|   76.50|comp
C0597249|ST|patf|    1.50|incp

C0597249|SG|DISO|  190.00|comp
# ------------------------------------------------------
```

Suggestions

Aligning SN and Meta relationships

- 54 types of SN relationships
- 106 additional types of Metathesaurus relationships

- Some are simply synonymous (*caused_by* / *due_to*; *follows* / *temporally_follows*)
- Some are specialized relationships (*manifestation_of* / *definitional_manifestation_of*)
- Many types of mapping relationships, not in SN

Add classification information to SN

- Explicit classificatory principles (in addition to textual definition and examples)
- Abandon economy principle and return to JEPD (jointly exhaustive/pairwise disjoint) approach

Metathesaurus editing environment

- Use SN/Meta relation consistency as a constraint for assigning semantic types
- Use SN relations to suggest labels for unspecified Meta relations
- Use SN/Meta relation consistency to guide the review by the Metathesaurus editors
  - Inaccurate categorization?
  - Inaccurate Metathesaurus relation?

Conclusions

Conclusions

- Simultaneously
  - Improve SN
  - Improve categorization

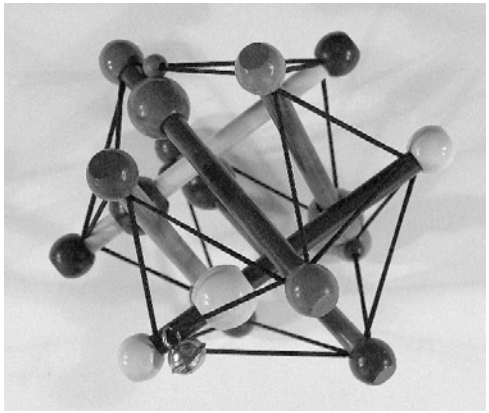
- ST assignment can be automated in part

Some references

- McCray AT, Bodenreider O.
  A conceptual framework for the biomedical domain.
  In: Green R, Bean CA, Myaeng SH, editors. The semantics of relationships: an interdisciplinary perspective. Boston: Kluwer Academic Publishers; 2002. p. 181-198.
- Bodenreider O, Burgun A.
  Aligning knowledge sources in the UMLS: Methods, quantitative results, and applications.
  Medinfo 2004:327-331.

Some references

- Burgun A, Bodenreider O.
  Aspects of the taxonomic relation in the biomedical domain.
  In: Welty C, Smith B, editors. Collected papers from the Second International Conference "Formal Ontology in Information Systems": ACM Press; 2001. p. 222-233.
- Mougin F, Bodenreider O.
  Approaches to eliminating cycles in the UMLS Metathesaurus: Naive vs. formal.
  Proceedings of AMIA Annual Symposium 2005:(submitted).



Medical Ontology Research

Contact: olivier@nlm.nih.gov

Web: mor.nlm.nih.gov

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA